**Stat and Data Analysis Project #2 – Predicting Body Fat**

***For this project you will be working in groups of 2.***

**INTRODUCTION:**

You will be given a Fathom document from a sample of adult males called **Body Fat**.

The data in this Fathom Document is from a study of adult males. The following variables were recorded for each male:

* Height(inches)
* Waist size(inches)
* Chest size(inches)
* Percent Body Fat(%bf) – calculated using a displacement test

**DIRECTIONS:**

You and your partner are to use Fathom, Excel, and Word to create a report (in Power Point) about the data. The report should have the following:

1. **Introduction**
2. **Analysis of Body Fat:**
   1. A histogram & boxplot of the percent body fat
   2. Summary Statistics (you should know which ones!)
   3. Test for outliers
   4. Description (shape, center, spread, anything unusual- use the correct measure of center & spread)
   5. What percent of males are overweight (over 19% body fat)? (show calculation)
   6. What **percentile** is the **first** male that is overweight? (show calculation)
   7. What percent of males are obese (over 25% body fat)? (show calculation)
   8. What **percentile** is the **first** male that is obese? (show calculation)
   9. What standard score (Z score) is the **first** male that is obese? (show calculation)
   10. How well does the distribution of percent of body fat seem to fit the normal model?
       1. What percent are within 1 standard deviation? (show calculation)
       2. What percent are within 2 standard deviations? (show calculation)
       3. What percent are within 3 standard deviations? (show calculation)
3. **Analysis of Height vs. Body Fat**
   1. Find the shortest **and** tallest males **and** what percent body fat **each** has.
      1. Does the shortest man have the lowest percent body fat?
      2. Does the tallest man have the highest percent body fat?
      3. Find the standard score for each man’s heights (show calculation)
      4. Find the standard scores for each man’s percent body fats (show calculation)
      5. Do these match up (the standard scores for the heights and the % body fats)? JUSTIFY.
4. **Analyzing how well each measurement is at predicting % Body Fat**

It is very costly and time consuming to measure a person’s body fat using a displacement test. Of the three measurements (height, waist size, and chest size) which can be used to create the best predictive model for the percent body fat for males? To analyze this, **do the following for EACH body measurement**.

1. Create a scatterplot with the measurement (X) vs. Body Fat (Y).
2. Describe the scatterplot (form, direction, strength, anything unusual)
3. Create LSRL on the plot (re-copy the scatterplot, but with the LSRL on it this time) and list the equation (with variable names in it)
4. Find correlation (r)
5. Interpret the slope
6. Find and interpret r2
7. Create residual plot & describe the form
8. Discuss whether the model is a good fit for the data (use residual plot, correlation, and the original scatterplot)
9. **Picking the best measurement (for predicting % Body Fat)**
   1. Decide the best model with good justification (out of the 3 measurements).
   2. Pick three men’s measurements (one low, one in the middle, one high) & predict their percent body fat. (show all work)
   3. Find the residual for each prediction. (show all work)
   4. List possible lurking/confounding variables for the model that you chose.
10. **Conclusion** (summarize your analysis)

**Rubric for Project #2**

**ITEM POINTS DEDUCTED**

1. **Introduction**  \_\_\_\_\_\_\_(5)
2. **Analysis of Body Fat:** 
   1. A histogram & boxplot of the percent body fat \_\_\_\_\_\_\_(6)
   2. Summary Statistics \_\_\_\_\_\_\_(4)
   3. Test for outliers \_\_\_\_\_\_\_(5)
   4. Description \_\_\_\_\_\_\_(5)
   5. What percent of males are overweight? \_\_\_\_\_\_\_(3)
   6. What percentile is the first male that is overweight? \_\_\_\_\_\_\_(3)
   7. What percent of males are obese? \_\_\_\_\_\_\_(3)
   8. What percentile is the first male that is obese? \_\_\_\_\_\_\_(3)
   9. What standard score (Z score) is the first male that is obese? \_\_\_\_\_\_\_(3)
   10. How well does the distribution of % body fat fit the normal model? \_\_\_\_\_\_\_(6)
       1. Percentages within 1, 2, 3 standard deviations?
3. **Analysis of Height vs. Body Fat**
   1. Shortest and tallest males and what percent body fat each has. \_\_\_\_\_\_\_(4)
      1. Does the shortest man have the lowest percent body fat? \_\_\_\_\_\_\_(1)
      2. Does the tallest man have the highest percent body fat? \_\_\_\_\_\_\_(1)
      3. Find the standard score for each man’s heights \_\_\_\_\_\_\_(4)
      4. Find the standard scores for each man’s percent body fats \_\_\_\_\_\_\_(4)
      5. Do these match up? JUSTIFY. \_\_\_\_\_\_\_(4)
4. **Analyzing how well each measurement is at predicting % Body Fat**

MEASUREMENT #1

* 1. Create scatterplot \_\_\_\_\_\_\_(3)
  2. Describe the scatterplot \_\_\_\_\_\_\_(4)
  3. Create LSRL on the plot & list equation (with variable names) \_\_\_\_\_\_\_(4)
  4. Find correlation (r) \_\_\_\_\_\_\_(1)
  5. Interpret the slope \_\_\_\_\_\_\_(3)
  6. Find and interpret r2 \_\_\_\_\_\_\_(3)
  7. Create residual plot & describe the form \_\_\_\_\_\_\_(5)
  8. Is the model is a good fit for the data? \_\_\_\_\_\_\_(4)

MEASUREMENT #2

1. Create scatterplot \_\_\_\_\_\_\_(3)
2. Describe the scatterplot \_\_\_\_\_\_\_(4)
3. Create LSRL on the plot & list equation (with variable names) \_\_\_\_\_\_\_(4)
4. Find correlation (r) \_\_\_\_\_\_\_(1)
5. Interpret the slope \_\_\_\_\_\_\_(3)
6. Find and interpret r2 \_\_\_\_\_\_\_(3)
7. Create residual plot & describe the form \_\_\_\_\_\_\_(5)
8. Is the model is a good fit for the data? \_\_\_\_\_\_\_(4)

MEASUREMENT #3

* 1. Create scatterplot \_\_\_\_\_\_\_(3)
  2. Describe the scatterplot \_\_\_\_\_\_\_(4)
  3. Create LSRL on the plot & list equation (with variable names) \_\_\_\_\_\_\_(4)
  4. Find correlation (r) \_\_\_\_\_\_\_(1)
  5. Interpret the slope \_\_\_\_\_\_\_(3)
  6. Find and interpret r2 \_\_\_\_\_\_\_(3)
  7. Create residual plot & describe the form \_\_\_\_\_\_\_(5)
  8. Is the model is a good fit for the data? \_\_\_\_\_\_\_(4)

1. **Picking the best measurement (for predicting % Body Fat)**
   1. Decide on best model with justification \_\_\_\_\_\_\_(4)
   2. Three men’s measurements (low, med, high) w/ prediction of % body fat \_\_\_\_\_\_\_(9)
   3. Residual for each prediction \_\_\_\_\_\_\_(6)
   4. Lurking/confounding variables \_\_\_\_\_\_\_(3)
2. **Conclusion** \_\_\_\_\_\_\_(8)
3. **Individual contributions** \_\_\_\_\_\_\_(25)
4. **Days late** \_\_\_\_\_\_\_(-21 per day)

**TOTAL \_\_\_\_\_\_\_/200**